

## 生成語彙論の枠組みに基づいた動的語彙モデルの構築

高橋 幸 李 相穆 茂木 和洋 小林 昌博 佐藤 滋†

† 東北大学大学院国際文化研究科 〒980-0871 宮城県仙台市青葉区川内

E-mail: †{sachi,sangmok,mgkz,mkoba,satos}@insc.tohoku.ac.jp

**あらまし** 我々は、生成語彙論 (the Generative Lexicon) [1] の枠組みに基づいた「動的語彙モデル」を提案する。このモデルは、辞書と規則体系から構成される。まず、生成語彙論の表示に従って辞書構築プログラムを開発し、名詞のタイプやイベントタイプによって、項構造と意味構造が自動的に対応するよう設計した。この辞書に登録された情報を利用して、限られた語彙から合成語や句を生成させるアルゴリズムを構築した。head となる単語に規則を適用し modifier の意味が加えられ、複雑な語構造が形成させる。また、単語の複数の意味は1つのエントリから派生され、文として構築していく際に、前後の単語の要求するタイプに従って、最も適当な意味が決定される。評価実験の結果から、このモデルが曖昧性の解消にも有効であることがわかった。

**キーワード** 動的語彙システム, 生成語彙論, 生成規則, 曖昧性の解消

### Building Dynamic Lexical Model Based on the Framework of the Generative Lexicon

Sachi TAKAHASHI, Sangmok LEE, Kazuhiro MOGI, Masahiro KOBAYASHI, Shigeru SATO†

† Graduate School of International Cultural Studies, Tohoku University

Kawauchi, Aoba-ku, Sendai-shi, Miyagi 980-8576 Japan

E-mail: †{sachi,sangmok,mgkz,mkoba,satos}@insc.tohoku.ac.jp

**Abstract** We propose a dynamic lexical model based on the *Generative Lexicon* [1]. This model is composed of a lexical database and a rule system. We first built the lexicon editor which allowed users to enter lexical items according to the *GL* feature structure format. Argument structure and qualia structure in each lexicon are automatically set by the categories of nominal types and event types. We built an algorithm to generate complex words or phrases from limited entries in this editor. Modifiers are added to a headed word to form complex word structures by tiered rule applications. In our model, multi-senses of a word are generated from a single lexical entry, and, when composing a sentence, the most appropriate sense is determined by type constraints which co-occurring words impose on the word. In an evaluation experiment, our dynamic lexical model was also proven to be suitable for resolving ambiguity.

**Key words** dynamic lexical system, the Generative Lexicon, generative rules, disambiguation

### 1. はじめに

辞書を構築するにあたり、複合語や派生語のような既存の語から語形成規則によって生成された合成語をどのように記述するかは重要な課題である。合成語は無限に造語されていくため、全てを辞書に登録することはできない。加えて、合成語には複数の意味解釈が可能な場合がある。そのため、1つ1つの語やその意味を数え上げによって辞書に記述することは、切りがなく大きな負荷となる。

そこで、語の持つ語彙情報から合成語を生成させ、また文脈によって1つの意味から多数の意味を動的に生成させるような語彙モデルを作る必要がある。そうすることは、辞書の冗長性を防ぎ、またその量の節約にも繋がる。

我々の提案する動的語彙モデルは、生成語彙論[1]の枠組みを用いており、図1のように辞書と規則体系からなる。単語の複数の意味は1つのエントリから派生され、文として組み上げ

られる際に、前後の単語の要求するタイプ制約に従って、適当な規則を適用させ、意味を生成させる。

我々は、まず、生成語彙論の表示に基づき、GUIの辞書構築プログラムを開発した。次章では、このプログラムについてその中で用いられるタイプ体系を中心に説明する。また、述語の意味を解析するには文の意味だけではなく、共起する名詞や副詞の情報が必要であることを例証を挙げながら考察する。第3節では、複雑な語構造を生成する規則を定式化し、その適用アルゴリズムを示す。第4節では、我々のモデルをコンピュータ上に実装化し、コーパスから抽出した文に対して、合成語の生成と曖昧性の解消が正確に行われるかどうか評価実験を行った結果を示す。最後に、本稿で提案したモデルに対する考察をまとめ、今後の課題を明らかにする。

### 2. 辞書

図2は生成語彙論の表示に基づき開発した辞書構築プログラ

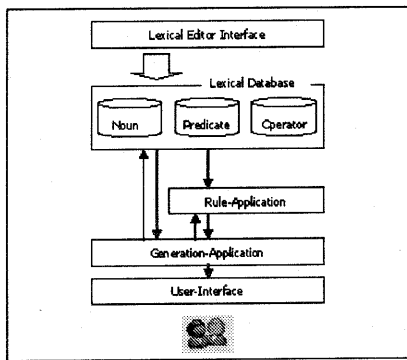


図1 動的語彙モデルの概要

ムの登録画面である。辞書登録の負担を減らすため、名詞のタイプや述語のイベントタイプによって分類し、そのタイプの選択により、項構造と意味構造が自動的に対応するよう設計した。また、IPAL [2][3][4] の辞書情報を読み込み、生成語彙論の枠組みにあうように自動的に加工し、人手によって修正を加えた。IPAL 辞書は名詞 1,081 語、形容詞 136 語、動詞 861 語から成る。この辞書は合成語を外し基幹語（単純語）のみで構成されており、少ない語彙から複雑な語を生成させるという我々の目的に添うものであったため、採用した。

我々のモデルでは語彙範疇を名詞、述語（動詞、形容詞、形容名詞）、副詞と格助詞等の機能語の3つに大別し、語彙範疇の選択によって各々の登録インターフェイスが切り替えられるよう設定した。副詞や付加詞は主辞事象の移動を促す operator、格助詞は項と述語を結び付ける operator として捉える。

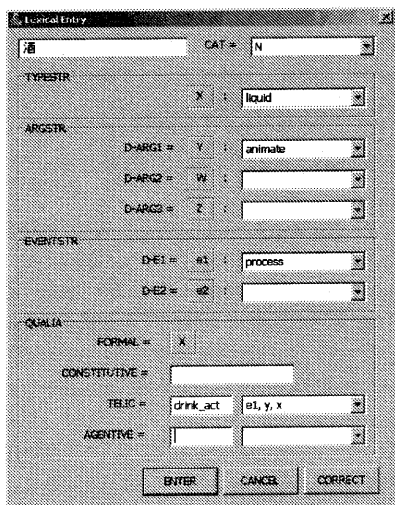


図2 辞書構築プログラムのインターフェイス (名詞版)

本節では、この辞書構築プログラムで用いた語彙のタイプ体系について説明する。

## 2.1 名詞の意味素性

生成語彙論では、述語のとる項を項構造 (ARGSTR) に記述する。各々の項はその名詞の属するタイプを与えられ、述語のとる項の制約となる。例えば、形容詞「悲しい」の主語は有生名詞である。「悲しい」の主語にあたる項には、animate の値を与えられる。実装化にあたり、我々は  $n$  個の項 (ARG<sub>n</sub>) に対して格 (主格), ヲ格 (対格), ニ格 (与格), 道具格等 (斜格) というように、斜格性の低いものから順に名詞を割り当てた。

また、名詞そのもののタイプを規定する。文献 [5] に基づき名詞のタイプを規定するのに、タイプ構造 (TYPESTR) を用いた。名詞に暗黙的に関わる項は D-ARG として項構造に記述し、対象物そのもののタイプはタイプ構造に記述される。「ナイフ」の場合、そのタイプ構造には人工的な道具であることから artifact.tool という値が、項構造には切る人を表す human やナイフで切る物体を表す phys\_obj の値が与えられる。

我々は生成語彙論の計算機構に合うように、IPAL による名詞の意味素性を修正した。分類とタイプ階層を図2に示す。このタイプ階層は、次節で述べる規則の適用条件の基盤となる。

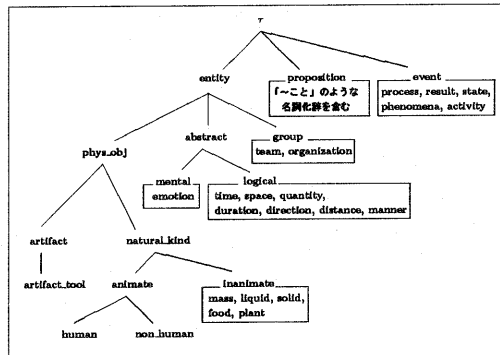


図3 名詞のタイプ階層

## 2.2 イベントタイプの分類

本節では、述語を (1) どのような項構造 (格パターン) を持つか、そして (2) どのような語彙的アスペクト構造を持つかの2つの観点から分類する。図4の手順に従って行った。

### STEP-1: 項構造の決定

日本語の述語はそのとる格パターンから他動詞、非能格自動詞、非対格自動詞に分類できる。他動詞はヲ格を伴い、その主語は意志の働く動作主 (Agent) である [6][7]。非能格自動詞と非対格自動詞の区別は、主語となるものが動作主であるか、それとも意志の働かない対象 (Theme) かの違いである。但し、表層的にはヲ格でも、深層的な意味では起点や終点を表すものが存在する。例えば、「仙台を出発する」の「仙台」は起点を表す。本稿では、主語の行為の対象をヲ格として扱う。まず、ヲ格の生起によって他動詞と自動詞を、次に非能格と非対格を区別した。また、変化動詞「なる」や存在動詞「いる」等、ニ格の生起によって必然的にとる述語についても分類した。その結果を表1に示す。

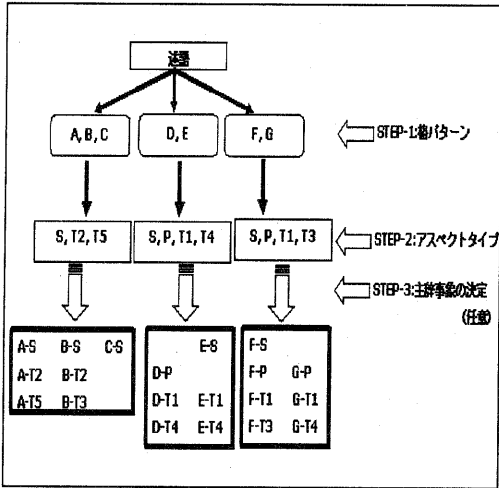


図4 イベントタイプの決定過程

表1 述語の種類別格パターン

分類	述語の種類	述語の例
A:ガ	非対格自動詞 (変化, 移動) 属性形容詞, 形容名詞	壊れる, 流れる, 高い, 豊か (な)
B:ガ, ニ (場所, 対象, 結果, 終点)	非対格自動詞 (状態, 変化)	ある, いる, なる, 変わる
C:ガ (主体, 対象)	感情形容詞	悲しい, 嬉しい, 悔しい
D:ガ	非能格自動詞 (活動)	働く, 走る, 急ぐ
E:ガ, ニ (場所, 対象, 終点)	非能格自動詞 (移動, 感情)	到着する, 驚く
F:ガ, ヲ	他動詞 (活動, 感情)	押す, 殺す, 食べる 愛する, 諦める
G:ガ, ヲ, ニ (結果, 終点)	他動詞	当てる, 変える, 合わせる

### STEP-2: 語彙的アスペクトの決定

次に、各々の述語が持つ語彙的アスペクトについて分類する。文献[8]では、図5のように英語動詞のアスペクトを3つの値 (state, process, transition) のいずれかを持つ1つ以上の事象とその時間関係、さらに主辞事象 (解釈される際の焦点になる事象) によって規定している。state は Vendler [9] の4分類の states, process は activities に対応し, transition は achievements と accomplishments を包括する。transition は下位イベントの組み合わせによって構成され (図5: E1 と E2), Vendler の accomplishments に対応する述語は結果状態が主辞事象となり, achievements の述語は結果に至るまでの過程が主辞事象となる。述語を S(tate), P(rocess), T(ransition) に3分類し, また各々の下位イベントを表2のように設定した。下位イベント process は主語が意志を持って働きかける過程, transition は無意識的に起こる変化を表す。

「 $e1 <_{\alpha} e2$ 」は  $e1$  が終了して, 次に  $e2$  が起こるという時間関係を示し, 「 $e1 <_{\alpha} e2$ 」は  $e1$  と  $e2$  が同時発生的に起こることを意味し, 「 $e1 <_{o_{\alpha}} e2$ 」は始めに  $e1$  が起こり, 次に  $e2$  が発生し, 両者が同じ終了時点まで継続するという関係を意味する。

### STEP-3: 主辞事象の決定

S, P のように下位イベントが存在しないもの, 「死ぬ」のように

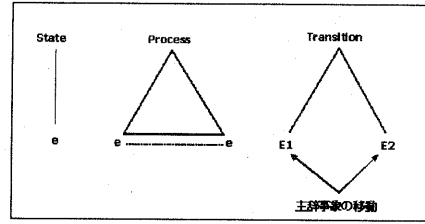


図5 文献[8]によるイベント構造の設定

表2 述語のアスペクトによる分類

アスペクトタイプ	下位イベント	述語の種類
S	state	非対格自動詞 (状態: ある, いる) 非能格自動詞 (感情: 壊れる) 他動詞 (感情: 信じる) 形容詞, 形容名詞
P	process	非能格自動詞 (活動: 踊る) 他動詞 (書く)
T1	process $<_{\alpha}$ state	非能格自動詞 (閉める) 他動詞 (建築する)
T2	transition $<_{\alpha}$ state	非対格自動詞 (変化: 暖れる)
T8	process $o_{\alpha}$ state	他動詞 (所有の移動: 買う)
T4	process $<_{o_{\alpha}}$ state	非能格自動詞 (移動: 歩く)
T5	transition $<_{o_{\alpha}}$ state	非対格自動詞 (変化: 高まる, 上がる, 腐る)

予め主辞事象が規定できるものに関しては、辞書に主辞事象を記載する。しかし、曖昧なものに関しては、underspecified (未指定) にしておき、文として組み上げた際に決定する。主辞事象の決定 operator としては、(a) 副詞, (b) 付加詞, (c) 名詞, (d) テンス・アスペクトが挙げられる。(1) に各々の例を示す。

- (1) a. 付加詞: 花子は、駅まで走った。
- b. 副詞: 太郎は 少しずつ 壁を白く塗った。
- c. 名詞: 彼は、住民票を仙台市に 移している。
- d. テンス・アスペクト: 煙が 上がっている。

(1a) では過去のある時点における結果状態まで含み, telic な意味解釈を持つ (「花子は走って駅に着いた」)。しかし, 「花子は、駅へ走った」の場合には、過去の process のみを表し, atelic な意味を持つ。このような解釈の違いは付加詞の種類によるものである。終点を表す「～まで」の場合には結果状態 state に焦点が置かれ, 「～へ」の場合にはそこに向かう process の方に焦点が置かれる。

(1b) の文は結果構文であるが, 進展性を計る副詞「少しずつ」が「壁を白く塗ること」の様態を修飾し, 塗る process の進展性に焦点が置かれる。形容詞のク形や形容名詞は process を修飾する副詞といえる。一方, 「壁を真っ白に塗る」の場合, 「真っ白に」が最上級を表す要素であるため, 進展性を計る副詞で修飾すると容認度が低くなる。「真っ白に」の場合には, 結果状態 state の方に焦点が置かれる。このように, 副詞によって修飾する下位イベントが異なる。「壁を30分間白く塗った」の「30分間」のように継続的な時間副詞の場合は process を修飾し, 「壁を30分で白く塗った」の「30分で」のように完結的な時間副詞の場合は白くなった状態ができあがったこと, つまり



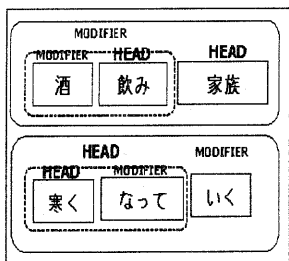


図6 語構造

の複合名詞のモデルを拡張し、日本語に適用した。図7にモデル化する。

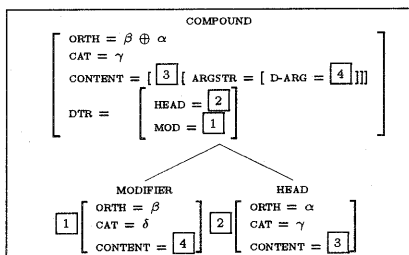


図7 複合規則

ORTH は正字上の並びを示し、COMPOUND の ORTH の値が  $\beta \oplus \alpha$  であれば、右側主要部の法則に従い  $\alpha$  が HEAD になる。

複合語の意味情報 CONTENT の値は、HEAD から受け継ぎ、MODIFIER の CONTENT の値が HEAD の D-ARG の値 [4] となる。HEAD の D-ARG の値のうち、どれと MODIFIER の値が一致するかは、(4) の2つのタイプ条件によって決まる。

- (4) a. HEAD の D-ARG のタイプと MODIFIER のタイプが一致する。  
 b. 図3の名詞のタイプ階層において、HEAD の D-ARG のタイプが MODIFIER のタイプより、上の要素である。

### 3.3 Transition 化規則

図7のように、述語の語幹にテンスやアスペクトや使役等を意味する接辞が付加することにより複雑な述語表現が形成される。こうした現象も一種の複合ということが出来る。

「走る」のような process を表す述語が「彼は駅まで走った」のように、telic な付加詞を入れることにより、「彼は駅まで走った、そして駅に着いた」という意味解釈が可能である。付加詞の追加によって、「走る」という process とその結果状態 state (「駅に着く」) の間に使役関係が生じる。「彼は駅まで走った」のイベント構造は、図8のように表すことができる。process は下位イベントをもたないが、終点を表す付加詞を含むと、下位イベント state が加えられ、全体として transition のイベント構造を構成する。また、「～なる」「～いく」のような述語につく接辞は、基となる述語のイベント構造に新たな transition 構造を加えることにより、その複雑な事象を説明できる。図9に「赤くなる」のイベント構造を表す。これを Transition 化と名づけた。Transition 化には (5) の条件が成り立つ。

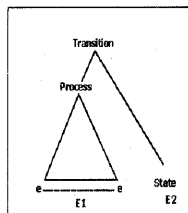


図8 「彼は駅まで走った」のイベント構成

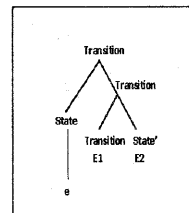


図9 「赤くなる」のイベント構成

- (5) Transition 化は、付加詞が使役関係の項となり得る場合、もしくは「～なる」や「～いく」などの状態変化を伴う接辞が付いた場合に起こり、最も右側の下位イベントが state になるまで続けられる。

### 3.4 意味拡張規則

意味拡張規則とは、デフォルトでデキゴトを表す名詞がモノ名詞になることを指す。意味拡張規則には3つのタイプが考えられる。

意味拡張規則1を図10に図式化する。この規則は(6c)のように人間や産物や道具の意味を表す複合語に対し適用する。モノ名詞の形成は、HEAD の D-ARG の値が COMPOUND の TYPESTR の値となることによって説明できる。

- (6) a. 昨日は酒を 飲み に行った。(行為)  
 b. 一晩中、酒飲み をした。(行為)  
 c. 私の家族は 酒飲み が多い。(人)

この規則で HEAD の D-ARG の値のうち、どれが COMPOUND の TYPESTR の値となるかは、(7) の条件で決められる。

- (7) a. 複合規則で MODIFIER と一致した D-ARG を除く。  
 b. 上以外の D-ARG で文中前後の単語の要求するタイプを持つ。

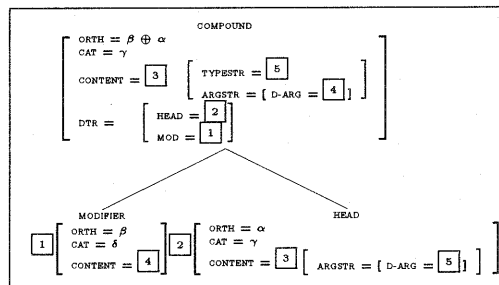


図10 意味拡張規則1

意味拡張規則2を図11に示す。この規則は場所、時間、状態の意味を表す複合語に対し適用する。文中、複合語の前後の単語がとる項のタイプとして、D-ARG<sub>n</sub> 内に適当なものがなければ、QUALIA の意味述語に關係する付加詞の項のタイプから導き出す。例えば、(8c)の「手洗い」は行為と場所を表す。

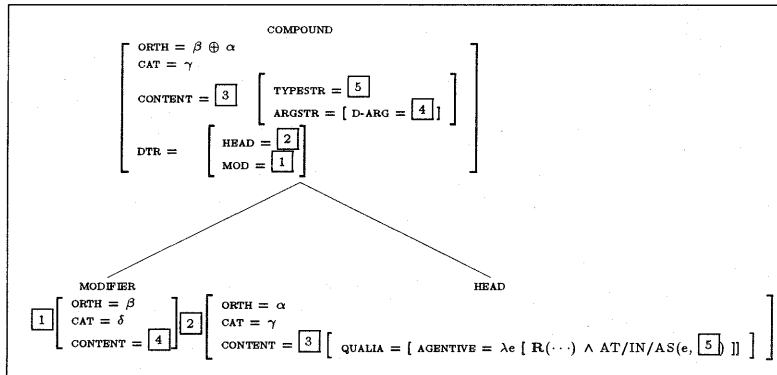


図 11 意味拡張規則 2

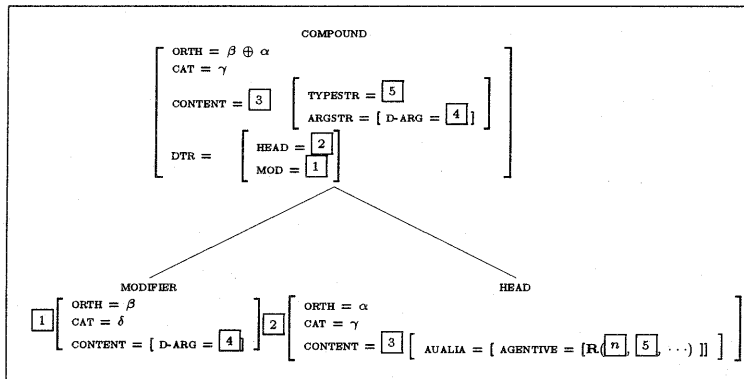


図 12 意味拡張規則 3

- (8) a. 洗いが悪い。(行為)  
 b. 外出後の手洗いは大切だ。(行為)  
 c. 手洗いはどこですか？(場所)

意味拡張規則 3 を図 12 に示す。(9b)の「木造建築」のように、MODIFIER が連体修飾として働いた結果、モノ名詞が生成される複合語に対し適用する。MODIFIER の D-ARG が HEAD の D-ARG と一致する。

- (9) a. 家の建築を断念する。(行為)  
 b. 日本の木造建築は貴重な財産だ。(行為)

以上の複合語生成規則を体系化したものを図 13 に示す。規則のうち、どの規則が適用されるかは、前後の単語の要求するタイプ制約によって決められる。この制約は、単語の意味を 1 つに決定すること、つまり曖昧性の解消にも役立つと考えられる。

## 4. 評価実験

### 4.1 実験方法

我々のモデルがより複雑な語構造の意味を適切に生成できるか、また文脈から判断して正しい意味解釈ができるか評価実験を行った。実験は図 14 の生成プログラムを用いて行った。対象データは『天声人語』1989 年の 1 年分と、CD-ROM 版『新潮文庫の 100 冊』(1995 年 新潮社)に収録の 5 作品から検索

した合成語を含む 500 文である。合成語の抽出にあたっては、形態素解析システム『茶筌』[12]を用いた。

IPAL の辞書情報を基に構築された辞書項目がリスト化され、抽出文を参考に単語を選択して文を形成する。実験データの中には登録されていない語も用いられているので、その都度登録する。但し、既に辞書の中に登録されている語を用いて作ることのできる合成語や句に関しては、図 13 の規則適用アルゴリズムによって語を生成できるよう設定した。

### 4.2 実験結果と考察

表 5 に生成実験の結果を示す(1 文中に複数の語形成のパターンを含むものがあつたため、合計は 500 語を超える)。約 78% の割合で正しく生成することができた。今回のデータでは意味拡張規則 3 で生成された複合語は少なかった。

曖昧性を解消する 1 つの手段として、前後の単語の要求する名詞のタイプによって単語の意味を決定させる手段が有効であった。図 3 のタイプ階層では名詞を大きく event, entity, proposition の 3 つに大別したが、曖昧性の解消にこの 3 タイプのどれを項として持つことができるか、次に他の項の意味素性との関連性を考慮して計算される。我々のモデルでは、ディフォルトの意味から、文脈による処理の結果意味が生成される過程を考察することができた。

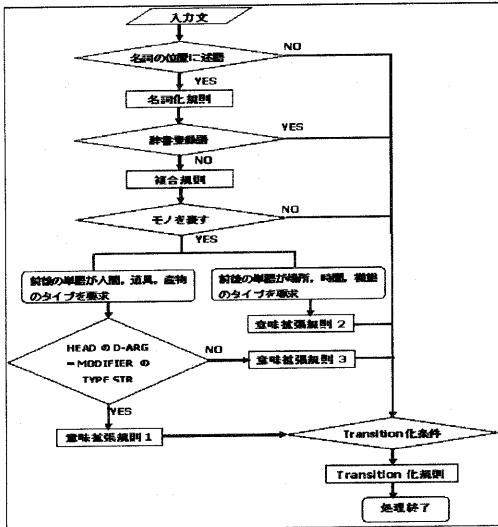


図 13 語形成規則適用アルゴリズム

表 5 生成実験の結果

語形成の種類	データ中に現れた数	正しく意味生成された数	割合
派生語	26	24	92.3%
複合語	368	364	98.9%
複合語 (意味拡張 1)	113	82	70.6%
複合語 (意味拡張 2)	56	41	73.2%
複合語 (意味拡張 3)	8	4	50.0%
述部	42	35	83.3%
合計	613	550	平均 78.0%

## 5. おわりに

本稿では、生成語彙論の枠組みを用いて辞書を構築し、それを利用して単語の意味解釈が環境によって動的に変化するモデルを紹介した。現在はまだ試験的な段階であり、生成プログラムを用いた実験は単文レベルでしか行っていない。また、今のところ生起制限がゆるいと思われる。しかし、素性構造を用いて単語の内部構造を捉えることができるため、統語的な関係の成り立たない複合語の生成を説明するのに本モデルは有効であると考えられる。今後は更に辞書エントリの数を増やすと共に、多くの事例を検証しながら規則体系をより強固なものにしていく予定である。

## 文 献

- [1] J. Pustejovsky, The Generative Lexicon, MIT Press, Cambridge, 1995.
- [2] 情報処理振興事業協会技術センター, 計算機用日本語基本動詞辞書 IPAL (Basic Verbs), 1987.
- [3] 情報処理振興事業協会技術センター, 計算機用日本語基本形容詞辞書 IPAL (Basic Adjectives), 1990.
- [4] 情報処理振興事業協会技術センター, 計算機用日本語基本名詞辞書 IPAL (Basic Nouns), 1996. (以上, <http://www.ipa.go.jp/STC/NIHONGO/IPAL/ipal.html> より入手)

- [5] M. Johnston and F. Busa, "Qualia structure and the compositional interpretation of compounds," *Breadth and Depth of Semantics Lexicons*, E. Viegas ed., Kluwer Academic Publisher, Dordrecht, 1999.
- [6] J. Grimshaw, *Argument Structure*, MIT Press, Cambridge, 1990.
- [7] 影山太郎, 動詞意味論, くろしお出版, 東京, 1996.
- [8] J. Pustejovsky, "The syntax of event structure," *Lexical and Conceptual Semantics*, B. Levin and S. Pinker eds., pp.47-81, Blackwell, Oxford, 1992.
- [9] Z. Vendler, "Verbs and times," *Philosophical Review*, 66, pp.143-160, 1957.
- [10] 高橋幸, 吉本啓, 佐藤滋, "構成要素の辞書情報を基にした複合名詞の意味形成制約," 言語処理学会第 8 回年次大会発表論文集, pp.160-163, 2002.
- [11] 影山太郎, 形態論と意味, くろしお出版, 東京, 1999.
- [12] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸 日本語形態素解析システム『茶筌』 version 2.2.9, 2002.